

AD-A159 131

NONPARAMETRIC METHODS FOR HAZARD RATE ESTIMATION FROM
RIGHT-CENSORED SAMP. (U) SOUTH CAROLINA UNIV COLUMBIA
DEPT OF MATHEMATICS AND STATISTI..

1/1

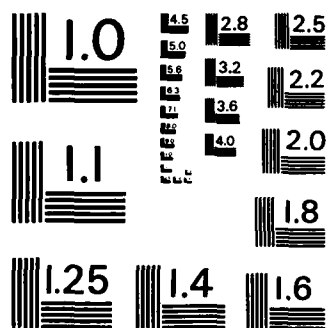
UNCLASSIFIED

D T MCNICHOLS ET AL. AUG 85 TR-106

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

AD-A159 131

NONPARAMETRIC METHODS FOR HAZARD RATE
ESTIMATION FROM RIGHT-CENSORED SAMPLES*

by

D. T. McNichols¹ and W. J. Padgett²

University of South Carolina
Statistics Technical Report No. 106

August, 1985

¹ Department of Statistics
Virginia Tech
Blacksburg, VA 24061

² Department of Statistics
University of South Carolina
Columbia, SC 29208

Approved for public release;
distribution unlimited.

DTIC
ELECTE
SEP 12 1985
S A D

*Supported by the U. S. Air Force Office of Scientific Research
under grant number AFOSR-84-0156 and the U. S. Army Research Office
under grant number MIPR ARO-139-85.

DTIC FILE COPY

ABSTRACT

Nonparametric estimation of the hazard rate of a lifetime distribution based on right-censored data is discussed. The methods considered include maximum likelihood, kernel-type, Bayesian, and histogram estimators. Recent results for kernel-type estimators are presented and compared.

Accession For	
WFO 00161	<input checked="" type="checkbox"/>
WFO 00161	<input checked="" type="checkbox"/>
WFO 00161	<input checked="" type="checkbox"/>
WFO 00161	<input checked="" type="checkbox"/>
by _____	
Classification _____	
Accession Number _____	
WFO 00161	
A-1	

AIR FORCE OFFICE OF SPECIAL INVESTIGATION
NOTICE
THIS
IS
A
CONFIDENTIAL
DOCUMENT
Chief, Research and Development Division



1. INTRODUCTION

Nonparametric estimation of the hazard rate or failure rate is a frequent topic of investigation in the statistical literature because of its practical importance. Until quite recently, hazard rate estimation had been based on complete samples of n independent identically distributed lifetimes. However, observations may be censored or truncated in many life testing situations (for example, see Lagakos, 1979).¹ This occurs often in medical trials when the patients may enter treatment at different times and then either die from the disease under investigation or leave the study before its conclusion. A similar situation may occur in industrial life testing when items are removed from the test at random times for various reasons. It is of interest to be able to estimate nonparametrically the unknown hazard rate of the lifetime random variable from this type of data without ignoring or discarding the right-censored information.

The purpose of this paper is to discuss nonparametric estimation of the hazard rate function for right-censored samples. The various types of estimators that have been proposed in the literature will be indicated and briefly discussed in Section 3. ^{These} ~~They~~ include maximum likelihood estimators, kernel type estimators, Bayesian estimators, and histogram estimators. Due to their computational simplicity and other properties, the kernel-type hazard rate estimators will be emphasized. Results of Tanner (1983) and Tanner and Wong (1983, 1984) will be presented in Section 4 while the estimator considered by McNichols and Padgett (1981) will be discussed in Section 5.

2. NOTATION AND PRELIMINARIES

Let X_1^0, \dots, X_n^0 be the survival times of n items or individuals which are censored from the right by a sequence of random variables U_1, \dots, U_n . Let F^0 denote the common unknown absolutely continuous distribution function of the X_i^0 's which are assumed to be independent and nonnegative, and let f^0 be a corresponding density function. The observed data will be denoted by the pairs (X_i, Δ_i) , $i=1, \dots, n$ where

$$X_i = \min\{X_i^0, U_i\}, \quad \Delta_i = \begin{cases} 1 & \text{if } X_i^0 \leq U_i \\ 0 & \text{if } X_i^0 > U_i. \end{cases}$$

Thus, it is known whether a particular observation is a time of failure (or death) or a loss time (censored time). The nature of the U_i 's determines the type of censoring. (i) If U_1, \dots, U_n are fixed constants, the observations are time truncated. If all U_i 's are equal to the same constant, then the case of Type I censoring results. (ii) If all $U_i = X_{(r)}^0$, the r th order statistic of X_1^0, \dots, X_n^0 , then the situation is that of Type II censoring. (iii) If U_1, \dots, U_n constitute a random sample from a distribution H (which is usually unknown) and are independent of X_1^0, \dots, X_n^0 , then (X_i, Δ_i) , $i=1, \dots, n$, is called a randomly censored sample. Assuming (iii), $\Delta_1, \dots, \Delta_n$ are independent Bernoulli random variables and the distribution function F of each X_i , $i=1, \dots, n$, is given by $1-F = (1-F^0)(1-H)$. If it is assumed that there is a positive constant β such that $1-H = (1-F^0)^\beta$, then this is called the Koziol-Green (1976) model of random censorship (which is the proportional hazards assumption of Cox, 1972). Chen, Hollander, and Langberg (1982) indicated that the pairs (X_i^0, U_i) , $i=1, \dots, n$, follow this

proportional hazards model if and only if (X_1, \dots, X_n) and $(\Delta_1, \dots, \Delta_n)$ are independent. This Koziol-Green model of random censorship arises in several situations (Efron, 1967; Csörgő and Horváth, 1981; Chen, Hollander and Langberg, 1982) and will be referred to in Section 5. In practice, however, this model may be somewhat restrictive.

Next, let Z_1, \dots, Z_n denote the ordered X_1, \dots, X_n with corresponding Δ -values, $\Delta_1, \dots, \Delta_n$. The Kaplan-Meier (1958) estimator $\hat{P}_n(t)$, or product-limit estimator (Efron, 1967), of the survival probability $1-F^0(x)$ at $x \geq 0$ is given by

$$\hat{P}_n(x) = \begin{cases} 1, & 0 \leq x \leq Z_1 \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1} \right)^{\Delta_i}, & x \in (Z_{k-1}, Z_k], k=2, \dots, n \\ 0, & x > Z_n. \end{cases}$$

Also, $\hat{F}_n(x) = 1 - \hat{P}_n(x)$ will denote the product-limit estimator of $F^0(x)$.

3. BRIEF REVIEW OF THE LITERATURE

The hazard rate function of the lifetime distribution F^0 is defined by $r^0(x) = f^0(x)/[1-F^0(x)]$, $x \geq 0$. Several authors have considered the non-parametric maximum likelihood estimation (MLE) of $r^0(x)$ assuming the random censorship model. Padgett and Wei (1980) obtained the MLE of r^0 in the class of increasing failure rate (IFR) distributions, while Mykytyn and Santner (1981) derived the MLE of r^0 assuming either an IFR, DFR, or U-shaped failure rate.

Bartoszyński, Brown, McBride and Thompson (1981) investigated the problem of estimating the intensity, or rate, function of a nonstationary Poisson

process arising in cancer studies under a type of right censorship. They considered a kernel estimator, a constrained maximum likelihood procedure, and a discrete maximum penalized likelihood approach.

Kernel-type estimators have been perhaps the most popular estimators in practice due to their relative computational simplicity, smoothness, and other properties. Blum and Susarla (1980) were the first to consider a kernel-type estimator of r^0 assuming randomly right-censored data. Some other kernel-type estimators for this case were obtained by Földes, Rejtő and Winter (1981), McNichols and Padgett (1981), Tanner (1983), Tanner and Wong (1983, 1984), and Schäfer (1985). These particular estimators will be discussed in more detail in Sections 4-6.

Burke (1981) considered a competing risk model (see also Burke and Horváth, 1982) in which the hazard rate estimator for the j th risk was either a kernel-type estimator or was of the form $r_{jn}(x) = \tilde{f}_{jn}(x)/(1-F_n(x))$, where $\tilde{f}_{jn}(x)$ was a kernel density estimator and F_n was the empirical distribution function of X_1, \dots, X_n .

Yandell (1983) obtained general delta sequence curve estimators and generalized the results of Watson and Leadbetter (1964) and Rice and Rosenblatt (1976) to the arbitrarily right-censored case. Using the martingale technique developed by Aalen (1978), Ramlau-Hansen (1983) considered the general multiplicative intensity model which included the kernel-type estimators of Földes, Rejtő and Winter (1981) and Yandell (1983).

Bayesian nonparametric estimators of the hazard rate function of the lifetime distribution have been obtained by Padgett and Wei (1981) and Dykstra and Laud (1981) for right-censored samples. Padgett and Wei (1981) used pure

jump processes as prior distributions on the hazard rate function which was assumed to be increasing, while Dykstra and Laud (1981) considered an extended gamma process as a prior. Lo (1978) also used a Bayesian nonparametric approach by constructing a prior random density as a convolution of a kernel function with the Dirichlet random probability. This technique could be applied to complete or censored samples.

In a recent paper Liu and Van Ryzin (1985) obtained a histogram estimator of the hazard rate function from randomly right-censored data and gave an efficiency comparison of their estimator with the kernel estimator of the hazard rate. Also, Liu and Van Ryzin (1984) gave the large sample theory for the normalized maximal deviation of a hazard rate estimator under random censoring which was based on a histogram estimate of the subsurvival density of the uncensored observations.

4. KERNEL-TYPE ESTIMATORS

Tanner and Wong (1983) considered the kernel-type estimator of r^0 given by

$$\hat{r}_n(x) = h^{-1} \sum_{j=1}^n (n-j+1)^{-1} \Delta_j' K\left(\frac{x-Z_j}{h}\right),$$

where K was a symmetric nonnegative kernel, $K(t) = o(t^{-1})$ as $t \rightarrow \infty$, and $\int_{-\infty}^{\infty} K(t)dt = 1$. The failure rate r^0 was assumed to be continuous and x satisfied $0 < F(x) < 1$. The following results for the mean and variance of $\hat{r}_n(x)$ were obtained:

$$E[\hat{r}_n(x)] = h^{-1} \int (1-F^n(y)) r^0(y) K\left(\frac{x-y}{h}\right) dy,$$

$$\begin{aligned} \text{Var}[\hat{r}_n(x)] = & h^{-2} \int I_n(F(y)) r^0(y) K^2\left(\frac{x-y}{h}\right) dy + 2h^{-2} \iint_{y \leq z} \{F^n(z) - F^n(y) F^n(z) \\ & - \frac{1-F(y)}{F(z)-F(y)} [F^n(z) - F^n(y)]\} r^0(y) r^0(z) K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) dy dz \end{aligned}$$

where

$$I_n(F) = \sum_{k=0}^{n-1} \binom{n-1}{k} F^k (1-F)^{n-k}.$$

In order to investigate the limiting behavior of these expressions the concept of compatibility was defined and is given below.

Definition. K is said to be compatible with a cdf F if for any $M > 0$, there exists h small enough such that $h^{-1} K(h^{-1}(y-x))/(1-F(y))$ is uniformly bounded for $|y-x| > M$.

Suppose the h_n 's satisfy $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Tanner and Wong (1983) then showed that $\hat{r}_n(x)$ was an asymptotically unbiased estimator of $r^0(x)$ whenever K was compatible with F^0 . Furthermore, if in addition K was also compatible with H , then

$$\text{Var}(\hat{r}_n(x)) = (nh)^{-1} \left(\int K^2(t) dt \right) r^0(x) (1-F(x))^{-1} + o((nh)^{-1}).$$

Thus, $\hat{r}_n(x)$ was a mean-square consistent estimator of $r^0(x)$ provided the aforementioned compatibility conditions held. In addition, by using Hajek's projection method, Tanner and Wong (1983) proved the asymptotic normality of $\hat{r}_n(x)$ (assuming K was compatible with F^0 and H).

Tanner and Wong (1984) also studied a class of estimators of the same general form as $\hat{r}_n(x)$ with $K(y/h)$ replaced by $K(y/\underline{\theta})$, where $\underline{\theta}$ was

a positive-valued "smoothing vector" chosen to maximize a likelihood function. Hence, for this estimator the smoothing parameters were chosen based on the observed data. Due to their complexity, these data-based estimators were studied by computer simulations.

Tanner (1983) considered a modified kernel-type estimator of the form

$$\hat{r}_n(x) = (2R_k)^{-1} \sum_{i=1}^n \frac{\Delta'_i}{n-i+1} K((x-Z_i)/2R_k),$$

where R_k was the distance from x to the k th nearest of the uncensored observations among X_1, \dots, X_n . This estimator allowed the data to play a role in determining the degree of smoothing that would occur in the estimate. Assuming that F^0 and f^0 were continuous in a neighborhood about x , $k = [n^\alpha]$, $\frac{1}{2} < \alpha < 1$, where $[\cdot]$ was the greatest integer function, that K had bounded variation and compact support on the interval $[-1, 1]$, and that r^0 was continuous at x , it was shown that $\hat{r}_n(x)$ was strongly consistent.

In a recent paper Schäfer (1985) considered a modified kernel-type estimator where the bandwidths were stochastic but did not depend on the point of interest x (as Tanner's 1983 estimator did). More specifically, Schäfer (1985) suggested the estimator

$$\hat{r}_n(x) = \sum_{j=1}^n (n-j+1)^{-1} \Delta'_j R_{n,j}^{-1} K\left(\frac{x-Z_j}{R_{n,j}}\right)$$

where

$$R_{n,j} = \inf \{r | H_n(Z_j+r) - H_n(Z_j-r) \geq p_n\},$$

p_n is a sequence of positive numbers for which $p_n \rightarrow 0$, and H_n is the

empirical cumulative hazard function introduced by Nelson (1969). Schäfer commented that the bandwidths suggested by Tanner (1983) had the disadvantage of being "biased" by the censoring distribution in the sense that they adapted to the conditional density of the random variables X_i^0 under the condition $X_i^0 < U_i$ of being uncensored, rather than to the hazard function r^0 to be estimated. Using the bandwidths suggested by Schäfer (1985), the uniform strong consistency of his estimator could be proven. More precisely, his result was the following.

Theorem. Choose p_n such that $p_n \log(n)/n^{1/2} \rightarrow \infty$. Let $K(\cdot)$ be a bounded Riemann integrable function with compact support on the interval $[-1,1]$. Let the density functions of both the X_i^0 and the U_i be continuous everywhere, and let $0 < b_1 < b_2$ with $P(U_i < b_2) < 1$. Then $\sup_{x \in [b_1, b_2]} |\hat{r}_n(x) - r^0(x)| \rightarrow 0$ with probability one.

5. KERNEL ESTIMATION UNDER THE PHM

In this section we will discuss the results of McNichols and Padgett (1981). Here the Koziol-Green model or proportional hazards model (PHM) defined in Section 2 will be assumed. Recall that this assumption is equivalent to assuming that $\underline{X} = (X_1, \dots, X_n)$ and $\underline{\Delta} = (\Delta_1, \dots, \Delta_n)$ are independent.

Although they are somewhat restrictive in practice, proportional hazards models have been used frequently in censored data problems to derive procedures and to study their theoretical properties. For the two sample test problem, Efron (1967) computed the Pitman efficiency of three competing test statistics under PHM. Koziol and Green (1976) derived the Crámer-von Mises statistic corresponding to the Kaplan-Meier product-limit estimate. Csörgő and Horváth

(1981) constructed empirical exact confidence bands for the life distribution of the censored variable and also gave rates of convergence for a wide class of functionals including the Crámer-von Mises functional. Chen, Hollander, and Langberg (1982) obtained small sample properties of the Kaplan-Meier estimator (in particular, the α th moment), while Cox (1972) derived procedures for regression models under a proportional hazards assumption. Such models also arise naturally due to the type of system under study in some reliability problems.

Throughout this section let $\tilde{r}_n(x)$ denote the hazard function estimator given by Watson and Leadbetter (1964a,b) for the uncensored case; that is, considering the Z_j 's as order statistics of a complete random sample from F ,

$$\tilde{r}_n(x) = h^{-1} \sum_{j=1}^n (n-j+1)^{-1} K\left(\frac{x-Z_j}{h}\right), \quad x \geq 0.$$

Let $K(t)$ be a suitable kernel and $h = h_n$ be a positive sequence. Based on the arbitrarily right-censored sample (X_i, Δ_i) , $i=1, \dots, n$, McNichols and Padgett (1981) considered

$$\bar{r}_n(x) = \left[\int_{-\infty}^{\infty} h^{-1} K\left(\frac{x-t}{h}\right) \frac{d\hat{F}_n(t)}{(1-\hat{F}_n(t))} \right] \cdot I_{[0, \infty)}(x)$$

as an estimator of the hazard function $r^0(x)$. This was understood to be

$$\bar{r}_n(x) = \left[h^{-1} \sum_{j=1}^n v_j K\left(\frac{x-Z_j}{h}\right) \right] \cdot I_{[0, \infty)}(x), \quad (5.1)$$

where $I_A(\cdot)$ is the indicator function of the set A ,

$$v_j = \frac{\hat{P}_n(Z_j) - \hat{P}_n(Z_{j+1})}{\hat{P}_n(Z_j)} = 1 - \left(\frac{n-j}{n-j+1}\right)^{\Delta_j'}, \quad j=1, \dots, n-1,$$

and $v_n = 1$.

The limiting behavior of (5.1) was considered under Parzen's (1962) conditions on the kernel function $K(t)$ and the sequence $\{h = h_n\}$ and Watson and Leadbetter's (1964a,b) conditions on K and F . These results are given in the next two theorems.

Theorem 5.1. Let $K(t)$ be a Borel function satisfying the conditions

$$(i) \quad \sup_{-\infty < t < \infty} |K(t)| < \infty,$$

$$(ii) \quad \int_{-\infty}^{\infty} |K(t)| dt < \infty,$$

$$(iii) \quad \lim_{t \rightarrow \infty} |t K(t)| = 0,$$

and

$$(iv) \quad \int_{-\infty}^{\infty} K(t) dt = 1.$$

Then for $x \geq 0$

$$E[\bar{r}_n(x)] = a E[\tilde{r}_n(x)] + (1-a)h^{-1} E_{Z_n} \left[K\left(\frac{x-Z_n}{h}\right) \right],$$

where $a \equiv P[X_1^0 \leq U_1]$. Furthermore, if

$$(i) \quad h = h_n > 0 \text{ is such that } \lim_{n \rightarrow \infty} h_n = 0,$$

$$(ii) \quad r(x) \text{ is continuous at } x \text{ (where } r(x) \text{ is the hazard function of the } X_1 \text{'s), and}$$

$$(iii) \quad K \text{ is compatible with } F,$$

then

$$\lim_{n \rightarrow \infty} E[\bar{r}_n(x)] = a r(x) = r^0(x), \quad x \geq 0.$$

Outline of Proof: Note that $\bar{r}_n(x)$ is a function of Z_1, \dots, Z_n and $\Delta'_1, \dots, \Delta'_{n-1}$. Also, since $\underline{\Delta}$ and \underline{X} are independent under PHM, \underline{Z} and $\underline{\Delta}'$ are independent. Thus,

$$\begin{aligned} E[\bar{r}_n(x)] &= E_{\underline{Z}, \underline{\Delta}'}[\bar{r}_n(x)] = h^{-1} \sum_{j=1}^n E_{\Delta'_j}(v_j) E_{Z_j} \left[K\left(\frac{x-Z_j}{h}\right) \right] \\ &= h^{-1} \sum_{j=1}^{n-1} E_{\Delta'_j}(v_j) E_{Z_j} \left[K\left(\frac{x-Z_j}{h}\right) \right] + h^{-1} E_{Z_n} \left[K\left(\frac{x-Z_n}{h}\right) \right] \\ &\quad + a h^{-1} E_{Z_n} \left[K\left(\frac{x-Z_n}{h}\right) \right]. \end{aligned} \quad (5.2)$$

But $E_{\Delta'_j}(v_j) = (n-j+1)^{-1} P[\Delta'_j = 1] + 0 \cdot P[\Delta'_j = 0] = a(n-j+1)^{-1}$,
for $j=1, \dots, n-1$ so that (5.2) becomes

$$a E[\tilde{r}_n(x)] + (1-a)h^{-1} E_{Z_n} \left[K\left(\frac{x-Z_n}{h}\right) \right] \rightarrow a \cdot r(x)$$

as $n \rightarrow \infty$ by Watson and Leadbetter's (1964a,b) results and the fact that

$$h^{-1} E_{Z_n} \left[K\left(\frac{x-Z_n}{h}\right) \right] \rightarrow 0.$$

By the PHM,

$$1-F(x) = [1-F^0(x)][1-H(x)] = [1-F^0(x)]^{1+\beta}.$$

Therefore,

$$f(x) = (1+\beta) f^0(x) [1-F^0(x)]^\beta,$$

and

$$a r(x) = \frac{a f(x)}{1-F(x)} = \frac{a(1+\beta) f^0(x) [1-F^0(x)]^\beta}{[1-F^0(x)]^{1+\beta}} = r^0(x),$$

since $a \equiv P[X_1^0 \leq U_1] = (1+\beta)^{-1}$ (Csörgő and Horváth, 1981). ///

Note that

$$\bar{r}_n(x) = \hat{r}_n(x) + (1 - \Delta'_n) h^{-1} K\left(\frac{x - Z_n}{h}\right), \quad (5.3)$$

where $\hat{r}_n(x)$ is the estimator of Tanner and Wong (1983). Thus $\bar{r}_n(x) = \hat{r}_n(x)$ whenever $\Delta'_n = 1$, that is, whenever the largest observation is uncensored. Also, (5.3) and the fact that $r(x) = r^0(x)$ under the PHM assumption account for the somewhat different expressions (for the mean and variance) obtained by McNichols and Padgett (1981) and Tanner and Wong (1983). The conditions on K assumed by Tanner and Wong were essentially Parzen's (1962) conditions. Also, under the PHM,

- (i) $r(x)$ is continuous implies that $r^0(x)$ is continuous
- and
- (ii) K is compatible with F implies K is compatible with F^0 .

The following result was proven for the second moment of $\bar{r}_n(x)$.

Theorem 5.2. Let $K(t)$ satisfy the conditions of Theorem 5.1. Then for $x \geq 0$,

$$E[\bar{r}_n^2(x)] = a^2 E[\tilde{r}_n(x)]^2 + (a-a^2)p_{1n} + (1-a)E_n + p_{3n},$$

where

$$p_{1n} = h^{-2} \sum_{j=1}^n \binom{n}{j-1} (n-j+1)^{-1} \int_0^\infty K^2\left(\frac{x-t}{h}\right) f(t) [1-F(t)]^{n-j} F(t)^{j-1} dt,$$

$$E_n = h^{-2} \int_0^\infty K^2\left(\frac{x-t}{h}\right) f(t) n F(t)^{n-1} dt = h^{-2} E_{Z_n} \left[K^2\left(\frac{x-Z_n}{h}\right) \right],$$

and

$$p_{3n} = 2(a-a^2)h^{-2} \int_0^\infty \int_0^{t_2} \sum_{i=1}^{n-1} \frac{n!}{(n-i+1)(i-1)!(n-i-1)!} F(t_1)^{i-1} [F(t_2) - F(t_1)]^{n-i-1} K_{f,n}(t_1, x) K_{f,n}(t_2, x) dt_1 dt_2,$$

with $K_{f,n}(t,x) = f(t)K((x-t)/h)$. In addition

$$\lim_{n \rightarrow \infty} E[\bar{r}_n^2(x)] = [\lim_{n \rightarrow \infty} E \bar{r}_n(x)]^2 = [r^0(x)]^2, \quad x > 0,$$

if $nh_n \rightarrow \infty$ and $n \cdot \sup_{y \geq \epsilon h_n} |y K(y)| \rightarrow 0$ as $n \rightarrow \infty$ for $\epsilon > 0$ fixed.

Corollary 5.3. Under the conditions of Theorem 5.2, for $x > 0$,

$$\lim_{n \rightarrow \infty} E[\bar{r}_n(x) - r^0(x)]^2 = 0.$$

Note that the last condition on K in Theorem 5.2 is satisfied for kernels with finite support and also for the normal kernel for choices of h_n such as $h_n = n^{-1/5}$. It should be noted also that $nh_n \rightarrow \infty$ is Parzen's (1962) condition. In addition, Theorem 5.2 implies that $\text{var}[\bar{r}_n(x)] \rightarrow 0$ as $n \rightarrow \infty$ for each $x > 0$.

Outline of Proof of Theorem 5.2: Squaring $\bar{r}_n(x)$ from (5.1) and taking the expectation,

$$\begin{aligned} E[\bar{r}_n^2(x)] &= h^{-2} \sum_{j=1}^n E_{\Delta}(v_j^2) E_{Z_j} [K^2(\frac{x-Z_j}{h})] \\ &\quad + 2h^{-2} \sum_{i < j}^n \sum_{i < j}^n E_{\Delta}(v_i v_j) E_{Z_i, Z_j} [K(\frac{x-Z_i}{h}) K(\frac{x-Z_j}{h})]. \end{aligned} \quad (5.4)$$

Now, the first term in (5.4) can be written as

$$\begin{aligned} ah^{-2} \sum_{j=1}^n (n-j+1)^{-2} E_{Z_j} [K^2(\frac{x-Z_j}{h})] &+ (1-a)h^{-2} E_{Z_n} [K^2(\frac{x-Z_n}{h})] \\ &\equiv a p_{1n} + (1-a)E_n \\ &= a^2 p_{1n} + (a-a^2)p_{1n} + (1-a)E_n. \end{aligned} \quad (5.5)$$

Next, for $i < j$, $j=2, \dots, n-1$,

$$E_{\Delta_i}(v_i v_j) = E_{\Delta_i}(v_i) E_{\Delta_j}(v_j) = a^{2(n-i+1)-1} (n-j+1)^{-1}$$

since v_i is a function only of Δ_i , and Δ_i and Δ_j are independent.

Also, $E_{\Delta_i}(v_i v_n) = E_{\Delta_i}(v_i) = a^{2(n-i+1)-1}$. Thus, the second term in (5.4) becomes

$$\begin{aligned} & 2h^{-2} \sum_{i < j}^{n-1} \sum_{j=2}^{n-1} E_{\Delta_i}(v_i v_j) E_{Z_i, Z_j} \left[K\left(\frac{x-Z_i}{h}\right) K\left(\frac{x-Z_j}{h}\right) \right] \\ & + 2h^{-2} \sum_{i=1}^{n-1} E_{\Delta_i}(v_i) E_{Z_i, Z_n} \left[K\left(\frac{x-Z_i}{h}\right) K\left(\frac{x-Z_n}{h}\right) \right] \\ & + 2h^{-2} \sum_{i=1}^{n-1} a^{2(n-i+1)-1} E_{Z_i, Z_n} \left[K\left(\frac{x-Z_i}{h}\right) K\left(\frac{x-Z_n}{h}\right) \right] \\ & = 2a^2 p_{2n} + p_{3n}, \end{aligned}$$

where

$$p_{2n} = h^{-2} \sum_{i < j}^n \sum_{j=2}^n (n-i+1)^{-1} (n-j+1)^{-1} E_{Z_i, Z_j} \left[K\left(\frac{x-Z_i}{h}\right) K\left(\frac{x-Z_j}{h}\right) \right]$$

and

$$p_{3n} = 2(a-a^2)h^{-2} \sum_{i=1}^{n-1} (n-i+1)^{-1} E_{Z_i, Z_n} \left[K\left(\frac{x-Z_i}{h}\right) K\left(\frac{x-Z_n}{h}\right) \right]. \quad (5.6)$$

Thus, from (5.4) - (5.6) we have

$$E[\tilde{r}_n^2(x)] = a^2 E[\tilde{r}_n(x)]^2 + (a-a^2)p_{1n} + (1-a)E_n + p_{3n} \quad (5.7)$$

since $p_{1n} + 2p_{2n}$ is simply $E[\tilde{r}_n(x)]^2$. This then implies that

$$\lim_{n \rightarrow \infty} E[\tilde{r}_n^2(x)] = a^2 r^2(x) = [r^0(x)]^2$$

since it can be shown that the p_{1n} 's, E_n 's and the p_{3n} 's converge to zero

as $n \rightarrow \infty$. Thus, under the PHM

$$E[\bar{r}_n^2(x)] \rightarrow [r^0(x)]^2. \quad ///$$

Note that Tanner and Wong (1983) required that K be compatible with both F^0 and H in order to derive their results for $\text{Var}[\hat{r}_n(x)]$. The compatibility condition of K with H is different from the supremum condition on K in Theorem 5.2.

6. CONCLUSION

Numerous methods for estimating the hazard rate of a lifetime distribution based on right-censored data are now available in the literature. These results are all fairly recent and appear to be growing rapidly. As the results in Sections 4 and 5 indicate, the kernel method has been most widely studied and perhaps provides the smoothest nonparametric estimates of the hazard rate. They are simple to compute for a given bandwidth h_n . An important problem which should be studied with respect to this estimator is how to optimally select the bandwidth for a given right-censored sample.

REFERENCES

- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist. 6, 701-726.
- Bartoszynski, R., Brown, B.W., McBride, C., and Thompson, J. R. (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. Ann. Statist. 9, 1050-1060.
- Blum, J.R. and Susarla, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. Multivariate Analysis-V, P.R. Krishniah, Ed., North Holland, NY, 213-222.
- Burke, M.D. (1981). Approximations of some hazard rate estimators in a competing risks model, in On Random Censorship: A Collection of Fifteen Papers, Carleton Mathematical Lecture Note No. 30, Carleton University, Ottawa.
- Burke, M. and Horváth, L. (1982). Density and failure rate estimation in a competing risks model. Preprint, Dept. of Math. and Statist., University of Calgary, Canada.
- Chen, Y. Y., Hollander, M., and Langberg, N.A. (1982). Small sample results for the Kaplan-Meier estimator. J. Amer. Statist. Assoc. 77, 141-144.
- Cox, D. R. (1972). Regression models and life-tables. J. Royal Statist. Soc. B34, 187-220.
- Csörgő, S. and Horváth, L. (1981). On the Koziol-Green model for random censorship. Biometrika 68, 391-401.
- Dykstra, R.L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. Ann. Statist. 9, 356-367.
- Efron, B. (1967). The two sample problem with censored data. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 4, 831-853.
- Földes, A., Rejtő, L., and Winter, B.B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data, Part II: Estimation of density and failure rate. Periodica Mathematica Hungarica 12, 15-29.
- Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.
- Koziol, J.A. and Green, S.B. (1976). A Cramér-von Mises statistic for randomly censored data. Biometrika 63, 465-473.
- Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.
- Liu, Y.C. and Van Ryzin, J. (1984). The asymptotic distribution of the normalized maximal deviation of a hazard rate estimator under random censoring. (To appear in Colloquia Mathematica Societatis Janos Bolyai, Debrecen, Hungary).

- Liu, Y.C. and Van Ryzin, J. (1985). Histogram estimation of failure rate and related functions under random censoring. Ann. Statist. 13, 592-605.
- Lo, A.Y. (1978). Bayesian nonparametric method for rate function. Dept. of Math. and Statist. Tech. Rep., University of Pittsburgh.
- McNichols, D.T. and Padgett, W.J. (1981). Hazard rate estimation under the Koziol-Green model of random censorship. Statistics Tech. Rep., University of South Carolina.
- Mykityn, S. and Santner, T.A. (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. Commun. Statist.-Theor. Meth. A10, 1369-1387.
- Nelson, W. (1969). Hazard plotting for incomplete failure data. J. Qual. Techn. 1, 27-52.
- Padgett, W.J. and Wei, L.J. (1980). Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations. Biometrika 67, 470-474.
- Padgett, W.J. and Wei, L.J. (1981). A Bayesian nonparametric estimator of survival probability assuming increasing failure rate. Commun. Statist.-Theor. Meth. A10, 49-63.
- Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. Ann. Statist. 11, 453-466.
- Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. Sankhya Ser.A 88, Pt. 1, 60-78.
- Schäfer, Helmut (1985). A note on data adaptive kernel estimation of the hazard and density function in the random censorship situation. Ann. Statist. 13, 818-820.
- Tanner, M.A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. Ann. Statist. 11, 994-998.
- Tanner, M.A. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. Ann. Statist. 11, 989-993.
- Tanner, M.A. and Wong, W.H. (1984). Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. J. Amer. Statist. Assoc. 79, 174-182.

Watson, G.S. and Leadbetter, M.R. (1964a). Hazard analysis I.
Biometrika 51, 175-184.

Watson, G.S. and Leadbetter, M.R. (1964b). Hazard analysis II.
Sankhya Series A 26, 110-116.

Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. Ann. Statist. 11, 1119-1135.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

7D-A159131

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Stat. Tech. Rep. No. 106			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-84-0156		
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research	
6c. ADDRESS (City, State and ZIP Code) University of South Carolina Columbia, SC 29208			7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR, ARO		8b. OFFICE SYMBOL (If applicable) NM		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0156	
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332			10. SOURCE OF FUNDING NOS.		
			PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2304	TASK NO. A5
11. TITLE (Include Security Classification) NONPARAMETRIC METHODS FOR HAZARD RATE ESTIMATION FROM RIGHT-CENSORED SAMPLES.					
12. PERSONAL AUTHOR(S) D. T. McNichols and W. J. Padgett					
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Yr., Mo., Day) August, 1985	
				15. PAGE COUNT 18	
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.	Random censorship; Kernel estimation; Failure rate estimation; Right-censoring; Incomplete data.		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>Nonparametric estimation of the hazard rate of a lifetime distribution based on right-censored data is discussed. The methods considered include maximum likelihood, kernel-type, Bayesian, and histogram estimators. Recent results for kernel-type estimators are presented and compared.</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff			22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5027		22c. OFFICE SYMBOL NM

END

FILMED

11-85

DTIC